The influence of short-term financial incentives on social norms and behaviors

January 2017

Jinhua Zhao John M. Kerr Maria Knight Lapinski Robert Shupp*

* Zhao (corresponding author): Department of Economics, Department of Agricultural, Food and Resource Economics, and Environmental Science and Policy Program, Michigan State University. Postal address: 486 W. Circle Drive, 110 Marshall Adams Hall, East Lansing, MI 48824. Voice: 517-353-9935. Email: jzhao@msu.edu.

Kerr: Department of Community Sustainability, Michigan State University, jkerr@msu.edu.

Lapinski: Department of Communication, Michigan State University, lapinsk3@msu.edu.

Shupp: Department of Agricultural, Food and Resource Economics, Michigan State University, shupprob@msu.edu.

Partial support for the project was provided by the Sustainable Michigan Endowment Project at Michigan State University, by the USDA National Institute of Food and Agriculture, Hatch project numbers MICL02244, MICL02173,and MICL02362, and by a National Science Foundation Award No. SMA-1328503.

Abstract

We develop and test a conceptual model on how financial incentives and social norms interact to influence pro-social behavior, in order to understand the long-term effects of short-term financial incentives in conditional payment programs. In a lab-based repeated public goods game, we find that although financial incentives raise contributes to the public good when they are in place, they reduce contributions after the payment stops. Specifically, subjects contribute less following the payment periods compared with those who do not receive payments at all, and the former even contribute less than before the payment was introduced. We distinguish between collective and perceived descriptive norms, and show that the latter's effect on contributions rises when subjects are more identified with their groups.

JEL: H41, C90, D03

Keywords: Public goods game; behavioral payment program; social norms; perceived descriptive norm; collective norm; group identification

1. Introduction

Since the late 1990s social programs worldwide have embraced the notion of offering financial incentives for desirable social behavior. Prominent examples of such *behavioral payment programs (BPP)* include payment for ecosystem services (PES), whereby land managers in environmentally sensitive areas receive direct payments to protect environmental amenities (Ferraro & Kiss, 2002; Wunder, 2005; Wunder, Pagiola & Engel, 2008), and conditional cash transfers (CCTs), in which impoverished households receive payment for human capital investments such as sending their children to school and taking them regularly to their local health clinic (Fiszbein & Schady 2009).

In this paper, we study the long-run effects of such payment programs when financial incentives are offered only in the short term. Most incentive-based social programs are funded by governments and donors and thus are vulnerable to budget tightening and changing priorities. Even in long-lasting programs, the budget for a given household or community may be short-lived. For example the Programa de Servicios Ambientales Hidrológicos in Mexico offers community-level contracts for five years that are potentially but not necessarily renewable, and the Costa Rican national PES program offers five-year afforestation contracts. Murillo et al. (2011) report that between 1997 and 2005, 500,000 ha in Costa Rica were covered by government-funded PES benefits but that by the end of 2006 only half as much area was still under PES.

The economic logic of BPPs is simple: pay people to undertake socially desirable behavior, and pay them enough to offset the opportunity costs of undertaking the behavior. This logic has little to say about what will happen once payment ends; economic theory predicts that behavior will return to its pre-payment state, unless the payment has contributed to changing the person's preferences or facilitated learning that enables a shift to a more socially preferable behavior. On the other hand, recent insights from behavioral economics suggest the potential for a more complex long-term effect of short-term payments. In particular, numerous studies have found evidence that monetary incentives can crowd out socially derived sources of motivation (Frey & Oberholzer-Gee 1997; Deci et al. 1999; Frey and Jengen 2001; Vollan 2008; Bowles & Polonia-Reyes 2011; Goeschl and Perino 2012; Rode et al. 2015), and that this effect can outlast the presence of the incentive. In one prominent study, Gneezy and Rusticini (2000) found that implementing a small fine for showing up late to collect children at day care centers actually increased the incidence of late arrivals, and that this increase persisted even after the fine was removed.

Building on this literature as well as recent developments in the communication sciences literature, we develop and test a conceptual model to explain how short-term incentives can affect pro-social behavior in the long run, emphasizing the role of social norms. The economics literature has posited a variety of reasons for the existence of crowding effects of monetary intervention (Nyborg and Rege 2003), such as altruism (Andreoni 1990), fairness (Fehr and Schmidt 1999), cognitive evaluation

(Frey & Oberholzer-Gee 1997), and social norms (Hollander 1990; Levitt and List 2007). However, these theories do not offer much help in explaining the long-term impacts of short-term financial intervention, as they have typically abstracted away from the evolution of these non-financial factors and how the evolution process is affected by financial intervention. Further, the literature has not explicitly studied how social norms interact with financial incentives in influencing behavior.[1]

The communication sciences literature, on the other hand, emphasizes the role of perceived social norms, and argues that the perceived norms can have direct as well as moderated impacts on behaviors (Lapinski & Rimal, 2005; Rimal, 2008). The literature distinguishes *collective norms,* which exist at the level of a community, from perceived norms, defined as perceptions about the prevalence or acceptability of a behavior (Lapinski & Rimal, 2005). Perceptions of prevalence of a behavior, in particular, are termed *perceived descriptive norms* (PDN: Cialdini, Reno, & Kallgren, 1990). In the present study, we also consider *collective descriptive norms* (CDN), defined as the observed prevalence of a certain behavior among members of a referent or social group. The theory of normative social behavior (TNSB), developed in Rimal & Real (2005), Lapinski & Rimal (2005), and Rimal & Lapinski (2015), identifies the moderators of the descriptive norm-behavior relationship where prior literature focused only on the main effects of social norms (c.f., Cialdini et al., 1990). It further argues that the influence of PDN on behaviors can be moderated by *group identification*, feelings of affinity with one's social group and the desire to be connected to that group. Thus, when one identifies more with her social group, the perceived behavior of fellow group members will have more influence on her own behavior so that the influence of PDNs will be heightened (Rimal & Real, 2005). When group identity is salient, individuals may experience positive affect when they behave normatively (Christensen et al., 2004), and individuals are aware that others in the group will endorse their compliance with those norms (Lapinski & Rimal, 2005).

Insights from the norms literature have been used to promote health behaviors (see Borsari & Carey, 2003; Mollen, Rimal, & Lapinski, 2010 for reviews). One example concerns binge drinking on college campuses, where students share strong group identity and many students perceive the rate of binge drinking among students to be much higher than it actually is. In this situation, use of norms messaging – a communication campaign – can be used to bring the PDN back in line with the CDN and help reduce the prevalence of binge drinking (DeJong & Smith, 2013). To do so requires knowledge of the PDN and CDN for a behavior about which there is agreement on its riskiness for members of the social system (Smith, Atkin, Martell, Allen, & Hembroff, 2006).

In this paper, we incorporate the more nuanced treatment of social norms from the communication literature to develop and test a theory on how BPP influences norms and subsequently

[1] Sethi & Somanathan (1996) examined the evolution of social norms but not in the context of motivational crowding out.

long-term behavior. At least theoretically, it is possible that the payment induces prosocial behavior and thus contributes to a descriptive norm of undertaking that behavior. On the other hand, if people know that the behavior is undertaken only for payment, the norms might be less influential on behavior. The relative effects of these competing forces on social norms will be determined by a variety of elements such as the program design and group identification.

We conduct a public goods lab experiment where the treatment groups receive incentives for public good contributions during some of the experiment rounds followed by rounds with no payments. This is similar to Kits et al (2014), who use a lab experiment to study contributions to an environmental amenity in which the subjects receive payments for their behavior. Our experiment includes a second treatment, in which subjects are grouped based on their homogeneity (in terms of propensity to contribute to the public good) to vary the degree and nature of norm formation. A novel contribution of our paper to the economics literature is that, in order to track how the norm evolves, we implement a set of measures of PDN and group identification through subject surveys.

The experiment results offer strong evidence in support of the tenets of our model. Consistent with economic theory, we find that payments significantly raise contributions to the public good, but only during rounds when such payments are made. The long-term effects of such payments are moderated by group identification and group homogeneity. Collective and perceived descriptive norms have different impacts on contribution behavior, and their effects operate through different channels. Specifically, the effects of PDN are moderated by group identification, while the effects of CDN are direct. While we emphasize the role of payment programs in this paper, we study the evolution of PDN and how it is affected by CDN and other factors in a companion paper (Lapinski et al. 2016),

The paper is organized as follows. In Section 2, we develop the theoretical model incorporating the communication treatment of social norms. We describe the experiment design in Section 3, analyze the experimental data in Section 4, and conclude in Section 5.

2. Social Norms and Decisions

In this section, we develop a theoretical model of decision-making in a public goods game that integrates communication theory on social norms with standard economic theory. Consider a public good contribution game among a group of N individuals indexed by $n=1, 2, \ldots, N$ over $t=1, \ldots, T$ periods. Each individual in each period is endowed with an initial income of w_n, which she can allocate between private consumption and contribution to the public good. Let $a_{n,t}$ be individual n's contribution to the public good in period t, and $g_t = \sum_{i=1}^{N} a_{n,t}$ be the level of public good provided, which is assumed for simplicity to equal to the aggregate contribution. A behavioural payment program (BPP) provides financial incentives to encourage prosocial behaviour, which in this case is contributions to the public good. When BPP is implemented, let P_t be the level of monetary incentives given to group members for each unit of their contribution. In this case, if individual *n* decides to contribute $a_{n,t}$, her private consumption becomes:

$$c_{n,t} \equiv w_n - (1 - P_t)a_{n,t} \tag{1}$$

An individual's utility consists of two components: a wealth component $U_n(\bullet)$ that is "neoclassical" in nature, derived from her consumption of private goods and the public good, and a social norm component $M_n(\bullet)$ that captures satisfaction (or costs) from conforming to (or deviating from) the group norm about contributions to the public good:

$$V_{n,t} = (1-\alpha_n)U_n(c_{n,t}, g_t) + \alpha_n M_n(a_{n,t}, D^p_{n,t}, D_{n,t}, G_{n,t}, \theta_{n,t}) \tag{2}$$

where $\alpha_n \in [0,1]$ is the weight attached to the norm component, $D^p_{n,t}$ measures n's *perceived* descriptive norm (PDN) in period *t*, $D_{n,t}$ measures the *collective* descriptive norm (CDN), $G_{n,t}$ is individual *n*'s identification or affiliation with the group, and $\theta_{n,t}$ is a parameter measuring the influences of the normative factors on her payoff. As we will show, the norm and group identification measures depend on what happened in previous periods and the dependence is further affected by the financial payments received in those periods. For example, if the individual believes that increased contributions in the previous period were brought forth by the payment program, she might be less responsive to the descriptive norms in making her own contribution decisions. In this case, $\theta_{n,t} = P_{t-1}$.

PDN is typically influenced by but could differ from the collective norm CDN, with the difference affected by communication messages and personal characteristics. Suppose an individual observes the group's average contribution level $\bar{a} = \sum_{n=1}^{N} a_n / N$. It is conceivable that the CDN is represented by $\bar{a}$. However, the individual might perceive that the group is less cooperative than reflected by $\bar{a}$ if she believes that external forces such as payment programs or government propaganda induced the individuals to contribute more. Her perceptions might also be influenced by processing biases and the social environment, especially for large and heterogeneous groups. In the binge drinking example, students typically look at social cues that are readily available to them and form perceptions that differ from the true prevalence of binge drinking, in most cases over-estimating its prevalence among college students.

The distinction between PDNs and CDNs is important for at least two reasons. First, if PDNs influence decisions over and above CDNs, the design of external interventions to promote pro-social behavior should attend to factors that influence perceptions. For example, financial incentives under BPPs might be amended with appropriate communication strategies to enhance not only CDNs but also PDNs. Second, PDNs and CDNs could influence behavior through different channels. Our experimental data show that while CDNs affect behavior directly, PDNs' influence on behavior is moderated by group identification: the influence is higher if the individual more strongly identifies with the group. This result implies that communication messages to enhance PDNs will be more effective if they are combined with interventions to enhance group identification.

In making her contribution decisions, an individual maximizes $V_{n,t}$ in (2) in each period, taking as given $D^{p}_{n,t}$, $D_{n,t}$, P_t, and the contribution levels of other individuals. The optimal conditions of the N individuals jointly determine the Nash equilibrium in period t. The equilibrium $\{a^{*}_{n,t}, n = 1, \ldots, N, t = 1, \ldots, T\}$ exists and is unique under the assumptions that (i) U_n is increasing and concave in c_n and g, and (ii) M_n is increasing and concave in a_n.

It is straightforward to establish that optimal contribution $a^{*}_{n,t}$ is increasing in payment P_t under the intuitive condition that $\partial U(\cdot)/\partial c > 0$ and $\partial^2 U(\cdot)/\partial c \partial g \geq 0$ (the private and public goods are weak compliments). It is also conceivable that a higher level of descriptive norm, perceived or collective, makes an individual's contribution more desirable: $\partial^2 M_{n,t} / \partial a_{n,t} \partial D^{p}_{n,t} > 0$, $\partial^2 M_{n,t} / \partial a_{n,t} \partial D_{n,t} > 0$. There could be multiple structural models or theories behind these conjectures. For instance, both the TNSB and the information cascade literature (Banerjee, 1992; Bikhchandani, Hirshleifer & Welch, 1992) imply that when one observes or perceives others contributing more, she might believe that the payoff from doing so must be higher, leading her to contribute more as well.[2] Lapinski et al. (2015) further shows that individual decisions can be positively affected by descriptive norms when simple heuristics are used in decision making. These arguments lead to the following testable hypothesis:

Hypothesis 1: An individual's contribution to the public good is increasing in (i) her collective descriptive norm and (ii) perceived descriptive norm.

[2] To formally establish this result, one needs to formally incorporate uncertainty into the payoff function and model learning among individuals. Our model in (2) is a reduced form where the normative influences are all incorporated into the term $M_n(\cdot)$.

The TNSB also argues that the effects of PDN are enhanced when the individual identifies more with the group: $\partial^3 M_{n,t} / \partial a_{n,t} \partial D_{n,t}^p \partial G_{n,t} > 0$. That is, an individual with greater group identification will be more responsive to the actions of other group members. The theory does not have predictions about whether the effects of CDN are also directly moderated by group identification.

Hypothesis 2: The effects of PDN on an individual's contribution are stronger if she identifies more with her group.

Economic theory suggests that at least in the short term, a BPP raises individual contributions: in period *t*, P_t unambiguously raises equilibrium contribution $a_{n,t}$. However, the presence of financial incentives might reduce the positive effects of CDN and PDN on contributions, as individuals might believe that increased contributions are due to the payment and thus carry less normative information. Thus, we expect that $\partial^3 M_{n,t} / \partial a_{n,t} \partial D_{n,t} \partial P_{n,t-1} < 0$ and $\partial^3 M_{n,t} / \partial a_{n,t} \partial D_{n,t}^p \partial P_{n,t-1} < 0$.

Hypothesis 3: (i) The direct effect of financial payments on individual contributions is positive, but (ii) the payment can reduce the positive effects of CDN and PDN on contributions.

Even if BPPs raise individual contributions in the short term, they may have unintended consequences in the long term after the program ends. In a companion paper (Lapinski et al. 2016), we investigate how the payments affect the evolution of individual PDNs through reducing the effects of such factors as the starting contribution levels of all group members. Presumably changes in PDNs after the BPP ends will continue to influence contributions. It is also possible that the negative effects on normative influences identified in Hypothesis 3 carry over to periods after the BPP ends, again leading to reduced contributions in the long term. Given data limitations, we will not be able to quantitatively identify the different channels through which the long-term effects materialize. Instead, we will test the following hypotheses about the overall long-term effects of BPPs:

Hypothesis 4: Behavioral payment programs can reduce contributions below their original level after the payment stops. Specifically, if payment is introduced in period τ given $2 \le \tau \le T-1$, then in period $\tau+1$, (i) those who have received the payment will contribute less than those who did not receive the payment in period τ, and (ii) the former will contribute less after the payment is over than before the payment was introduced:

$$a_{n,\tau+1} \mid n \text{ received payment in } \tau \quad < \quad a_{m,\tau+1} \mid m \text{ did not receive payment in } \tau;$$
$$a_{n,\tau+1} \mid n \text{ received payment in } \tau \quad < \quad a_{n,\tau-1} \mid n \text{ received payment in } \tau \tag{3}$$

Some groups might be more susceptible to the negative long-term effects of short-term payments than others, depending on the group structure and the strength of their norms. Conceivably, a group is less susceptible if its norms are strong enough to overcome the negative influences of the payments. In our experiment, we will not be able to directly manipulate the norm structure of the groups. Instead, we will sort subjects into groups based on their initial contributions to the public good, with the aim of creating relatively homogeneous groups with different starting norms about contributing to the public good. We will evaluate how the "sorted" groups differ from the unsorted groups, and among the former, how groups with higher initial contributions differ from those with lower initial contributions.

Hypothesis 5: The long-term effects of BPPs differ between homogeneous and heterogeneous groups.

3. Experiment Design and Estimation Model

To test how contribution decisions are affected by the different kinds of norms, how these factors are in turn affected by BPPs, and ultimately how BPPs affect long-run contribution after the payments are over, we conduct a computer-based public goods experiment with undergraduate student subjects at a major US university using Z-tree (Fischbacher 2007). Each subject is given an initial allocation of tokens, and makes decisions about investing the tokens into two accounts: 1) a public account whose benefits are shared by all group members, and 2) a private account that benefits the individual only. The payoff structure is set up so that aggregate payoff is maximized if subjects invest all tokens in the public account, but a selfish individual has an incentive to invest only in the private account (Fishbacher and Gachter 2010). We adopt a standard linear experimental design, where the monetary payoff is given by

$$U_{n,t} = w_n - a_{n,t} + 0.4g_t = w_n - a_{n,t} + 0.4\sum_{i=1}^{N} a_{i,t} \quad . \tag{4}$$

Specifically, each group includes four subjects, i.e., $N = 4$, and each subject is allocated 20 tokens, i.e., $w_n = 20$. An investment of one token in the private account yields a return of 1, and an investment in the public account yields a return of 1.6 that is divided among the four players such that each individual gets 0.4. Each token is worth \$0.04 and each subject's earnings ranged from about \$23 to \$38.

A total of 192 subjects are divided into 10 sessions, receiving one of four treatments to be discussed below (Table 1). Figure 1 illustrates the experiment procedure. Each session includes nineteen *periods*, indexed by $t = 0,1,...,18$. Contribution levels of period zero are used to sort and group two thirds of the subjects, while the remaining $t = 1,...,18$ periods are divided into three *phases*, with each phase consisting of six *rounds*. That is, *periods* are relative to the entire session and vary from 0 to 18, while *rounds* are specific to a particular phase and vary from 1 to 6. At the beginning of phase 1, after the groups (of four members each) are formed, each subject is informed of what each of her fellow group members contributed in period 0 and that she will be playing the game with the same group members for the entire phase of six rounds, but is not informed of whether or not there will be a next phase. Once the game moves to phase 2 (or 3), she is informed that her group remains intact and she will play the game with them for another six rounds, but not that there will be a next phase (or that the game will end). At the end of each round, each subject learns the group's total contribution to the public good in this round and her earnings from this round. At the end of each phase, each subject is asked a set of perception questions used to measure her PDN and group identification (Table 2). Such perceptions form the starting belief for the next phase. For phase one, the perception questions are also asked after round one, measuring the starting belief for rounds 2 to 6 in phase one.

The perception questions in Table 2, which are self-report items measuring PDN and group identification (GI), are derived from research presenting evidence for scale construct validity and reliability (e.g. Lapinski et al., 2013; Rimal & Real, 2005). PDN measures are assessed through four Likert-type items drawn from Lapinski et al. (2013) with 5-point response scales in which higher scores indicate greater prevalence perceptions. These multiple item measures are then subjected to confirmatory factor analysis (CFA) procedures (e.g., Hunter & Gerbing, 1982), which extend correlation analysis to test the extent to which these items are similar in measuring the underlying latent construct (PDN in this case). CFA applied to our data indicates strong construct validity of the PDN scale across the administration periods (i.e., periods when the survey questions are asked); thus, the items can be summed to form a unidimensional scale. Cronbach's alpha testing indicates strong evidence for scale reliability (with reliability coefficients ranging from $\alpha = 0.90 - 0.94$ across administration of the instrument). Group identification is measured using four indicators of perceived similarity with group members with response scales ranging from 1-7 with higher scores indicating greater similarity (Rimal, 2008). Again, CFA applied to our data indicates that the items could be summed together in a unidimensional scale, with reliability coefficients ranging $\alpha = .86 - .91$ across the three administrations of the measures.[3]

We use a 2x2 design with two treatments, "payment" and "sort" (Table 1). Subjects are assigned to sessions, which are randomly allocated to the four cells in Table 1. In about half of the sessions, subjects receive a "payment treatment" in which payment out of the public account is subsidized at a rate

[3]Additional details of the measurement analysis are available upon request.

of 0.6 *during phase 2*, so that (4) becomes $U_{n,t} = w_n - a_{n,t} + (0.4+0.6)g_t = w_n - a_{n,t} + \sum_{i=1}^{4} a_{i,t}$ for $t = 7,\ldots,12$. With this payoff structure, contribution to the public account weakly dominates contribution to the private account. Under the "sort treatment," received by about two thirds of the subjects, subjects in the treated session (usually consisting of 20-24 subjects in 5-6 groups) are ranked from high to low in terms of their contributions in period $t = 0$ and are then divided into four member groups from the most cooperative to the least cooperative. That is, the top four contributors form one (most cooperative) group, followed by the next four contributors, and so on. Among the resulting sorted groups, the top two represent "cooperative groups" and are labeled "sort-high" in our analysis, the next two represent "somewhat cooperative groups" and are labeled "sort-medium." The remaining sorted groups represent "uncooperative groups" and are labeled "sort-low." Recall that at the beginning of phase 1, a subject learns the contribution level of each of her fellow group members in period 0, forming initial beliefs about the group. This belief is then influenced by her observations of the aggregate group contribution levels in the subsequent rounds.

By grouping together subjects with similar inclinations towards public good contribution, we hope to test whether norms in homogeneous groups and especially in homogeneous and cooperative groups can better sustain the intervention of financial payments (Hypothesis 5). It has been widely observed that public good and common pool resource games played in laboratories exhibit rapid decline in cooperation over rounds of the game. However, Gachter and Thoni (2005) and Gunthorsdottir et al. (2007) found that cooperators who were grouped together maintained cooperation at a much higher level throughout the game. One possible explanation is that homogeneous cooperators might be able to maintain their "cooperative norm" while such norm is destroyed in other groups faster. If this is true, then these groups might also be more effective in defending their norms against financial payments. The "sort" treatment allows us to study whether this hypothesis is supported by our data.

The estimation equation for contribution a_{nkrj} of individual *n* of group *k* in Round *r* of phase *j* is

$$a_{nkrj} = \beta_0 + a_{n,0}\beta_1 + g_{k,0}\beta_2 + \text{ Norms } \beta_3 + \text{ Payment } \beta_4 + \text{Interactions } \beta_5 + \text{Z}_{nkrj}\beta_6 + \varepsilon_{nkrj} \qquad (5)$$

In (5), $a_{n,0}$ is individual n's contribution level in period $t = 0$, a summary measure of her inclination towards contributing to the public good. $g_{k,0}$ is the average contribution of all four members of group k in period 0, representing the overall inclination of the group towards contributing to the public good. "Norms" is a collection of norm-related variables, including CDN_{nkrj}, PDN_{nkrj}, and GI_{nkrj}. CDN is measured by the total contribution of the fellow group members in the previous period:

$CDN_{nkrj} = \sum_{i \neq n} a_{ik(r-1)j}$ for $r = 2, \ldots, 6$, and $CDN_{nk1j} = \sum_{i \neq n} a_{ik6(j-1)}$ for r =1. As discussed earlier, measures of PDN and GI are taken at the end of each phase and at the end of Round 1 of phase 1 (Figure 1), and become the starting perception values for the immediate next round. Thus, within phase 1, the starting perception measure exists for Round 2, while within phases 2 and 3, the starting perception measures exist for Rounds 1. We use linear smoothing to obtain the perception measures for the in-between rounds. For example, in the case of PDN,

$$\begin{aligned} PDN_{nkr1} &= \rho_{rj} PDN^{e}_{nk11} + (1-\rho_{rj}) PDN^{e}_{nk61}, \quad \rho_{rj} = (7-r)/5, \quad \text{for } r = 2, \ldots, 6 \\ PDN_{nkrj} &= \rho_{rj} PDN^{e}_{nk6(j-1)} + (1-\rho_{rj}) PDN^{e}_{nk6j}, \quad \rho_{rj} = (7-r)/6, \quad \text{for } r = 1, \ldots, 6;\ j = 2, 3 \end{aligned} \tag{6}$$

where PDN^{e}_{nkrj} represents the PDN value measured at the *end* of Round *r*. The first equation in (6) describes linear smoothing for phase 1. The PDN for round 2 (i.e., $r = 2$) equals PDN^{e}_{nk11}, the value measured at the end of round 1, while the PDN for other rounds are the weighted averages of the PDN values measured at the end of rounds 1 and 6, with the weights varying linearly in r. The second line equation in (6) describes the linear smoothing for phases 2 and 3. The GI measures are developed similarly.

In (5), "Payment" includes two dummy variables: $Payment_{kj} = 1$ if group k received a subsidy in phase j for contributing to the public good and $= 0$ otherwise, and $Post\text{-}Payment_{kj} = 1$ if group k received a subsidy in the *previous phase* $j-1$, and $= 0$ otherwise. "Interactions" include a set of interaction terms of variables from "Norms" and "Payment," used to study channels through which norms and payments affect the contribution behavior. Z_{nkrj} is a set of standard controls, including dummy variables for phases, rounds, and lagged contributions to control for unobserved individual characteristics.

Finally, ϵ_{nkrj} is the unobserved error term that is assumed to be *iid* across groups, rounds and phases. We allow the error term to be correlated across players within the same group, due to such reasons as strategic interactions and unobserved group characteristics, and estimate (5) using OLS with group clustered robust standard errors. Presumably one could explore the panel structure of the data, e.g., using a random effect estimator. We choose not to do so for two reasons. First, initial contributions $a_{n,0}$ and $g_{k,0}$ are invariant across experiment rounds, implying that the OLS and random effect estimators are identical. Second, even if the initial contributions are not included in (5), the efficiency gain of the random effect estimator is minimal since the group size in our experiment is small [see Cameron and Miller (2015) page 326 for details].[4]

4. Data and Estimation Results

Figure 2 presents the average individual contributions in each period for each treatment, and Table 3 shows the contributions in period 0 (the sorting period) as well as those averaged over the three phases. Period zero contributions are not statistically different across the payment vs. no-payment and sorted vs. not-sorted treatments (Wilcoxon Signed Rank test), indicating that the subjects are randomly assigned to the treatments. For phases without payments, although the unique Nash equilibrium predicts zero contributions, the average contributions exceed zero, consistent with typical findings in public good games. Although payment raises contributions, the average contribution is still less than 20. Finally, consistent with the literature, contribution levels fall over rounds in each phase and there is a restart effect as a new phase begins.

Payment significantly raises the contribution levels during phase 2 relative to groups without payment, as indicated in Figure 2 and by comparing the phase 2 average contributions with and without payments in Table 3. The increase is the largest for the sort-low (the least cooperative) groups (from 2.06 to 15.92). For sort-high and sort-medium groups, payment also mitigates the reduction of average contributions from phases 1 to 3. Specifically, the payment treatment limited the reduction to 8% and 6% relative to 15% and 19% without the payment treatment. However, the sort-low group that received payment, saw a reduction of 33% compared to the 15% reduction for the sort-low group without payment. Overall, these comparisons show modestly positive long-term effects of payment within the sorted group treatments. Note though that for the not sorted groups payment does not seem to have a long-term impact, either positive or negative. We now turn to the more comprehensive multivariate analysis.

Table 4 contains the explanations and summary statistics of the variables used in the regression analysis, and Table 5 shows the least square estimates and the clustered (by group) robust standard errors for different variations of the basic specification in (5). In Model (I), only initial contributions and payments are included as explanatory variables. Model (II) adds normative variables (CDN, PDN and GI), and their interactions are included in Model (III). The interactions of the normative variables and payments are introduced in Model (IV), and Model (V) includes the sorting treatment.[5] In all specifications, we include dummies for Rounds $r = 2,\dots,6$ and phases $j = 2,3$, in order to control for

[4] We did estimate a random effects model with the initial contributions excluded. The resulting estimates are indeed close to the OLS estimates.

[5] Since sorting is based on period 0 contributions, we remove “Period 0 group average contribution” in Model (V). The estimates are similar when this variable is included in the regression.

the trend of decreasing contributions as the game moves closer to the last period. The presence of time trends also suggests that, since individuals have acted dynamically in making their contribution decisions, there might be unobserved dynamic linkages across periods even after we control for the normative and payment variables. We include the lagged dependent variable (LDV), i.e., an individual's contribution in the previous period, to control for such dynamic considerations. We show later (in Table 6) that our main results are robust to this specification. We also report the marginal effects of the payment and norm variables at the bottom of Table 4, evaluated at sample averages. Since *Payment* and *Post-Payment* are dummy variables, their marginal effects are measured as the changes in average individual contributions when either variable's value increases from zero to one.

Across almost all specifications, an individual's contribution is increasing in her *Period 0 contribution* $a_{n,0}$ – its coefficient is significant at the 10% level. An individual's initial contribution is thus a good indicator of her "inclination" to contribute to the public good, lending support for relying on $a_{n,0}$ to sort the groups. The *Period 0 group average contribution* $g_{k,0}$ is positive and statistically significant in Model (I), but it is no longer significant once other group-related variables such as CDN and Payment are included. Since CDN is measured as the average group contribution in the previous period, the effects of $g_{k,0}$ are likely to be partly captured by CDN.

The effects of normative variables

Our results show that CDN and PDN have different impacts on contribution behavior. A higher CDN significantly raises the contribution level and this effect is robust across all specifications. The coefficient of PDN is not statistically significant by itself in Model (II), and becomes significant only when interacted with other variables. Across all samples, the average effect of PDN on individual contributions is not significant, while its effect is positive and significant for groups with high GI and negative and significant for groups with low GI. Partly confirming Hypothesis 1, the marginal effects of CDN are positive and significant across all models, while those of PDN are statistically not significant.

Models (III) – (V) show that the effects of CDN and PDN manifest through different channels. The effect of PDN is moderated by GI: the coefficient of $PDN*GI$ is positive and significant, implying that PDN positively affects an individual's contribution to a higher degree if she identifies more with her group. In contrast, the coefficient of $CDN*GI$ is not significant, implying that, unlike PDN, the effect of CDN is not moderated by GI. This finding confirms Hypothesis 2 and is consistent with our theory. These findings highlight the empirical relevance and importance of distinguishing between CDN and PDN.

The effects of CDN and PDN are influenced by the contribution subsidy differently. In Models (IV) and (V), we include interaction terms of CDN/PDN with the payment dummy $Payment_{kj}$, which equals one when individuals in group *k* receive the subsidy in (all rounds of) phase *j* and zero otherwise, and the post-payment dummy $Post\text{-}Payment_{k(j-1)}$, which equals one if group *k* individuals in phase j received subsidy during the previous phase $j-1$, and zero otherwise. The impacts of CDN on contribution levels are reduced by the presence of payment subsidies – the coefficient of $CDN * Payment$ is negative and statistically significant. Knowing that the higher contribution levels are partly incentivized by the contribution subsidy reduces the positive response of one's own contribution to CDN. This result confirms Hypothesis 3(ii) regarding CDN and is consistent with the interpretation that the positive effects of CDN might be due to reciprocity, where the contribution subsidy mitigates the need for "rewarding others' good deeds." In contrast, we find no evidence to support Hypothesis 3(ii) regarding PDN: the interaction term of Payment with PDN is not statistically significant, indicating that the effects of PDN are not influenced by the contribution subsidy. Note that the interaction terms of Post-Payment with CDN and PDN are not significant, indicating that behavioral payments moderate the effects of normative variables only when the payment is in place.

The effects of contribution subsidies

Confirming Hypothesis 3(i), Payment raises individual contributions: the coefficient of Payment is positive and statistically significant across all specifications in Table 5. Its marginal effect is positive and significant, and does not vary significantly across the model specifications. Note that since Payment is introduced only in phase 2, and we already include a phase 2 dummy in the regressions, the marginal effect of Payment measures the increase in average individual contributions *in phase 2* for groups that received the payment relative to those without the payment.

The coefficient of Post-Payment is negative and significant in Models (II) and (III) but is no longer significant when its interaction with other variables is included. Its marginal effect is negative and statistically significant. Again, since Post-Payment equals one only in phase 3 (for groups that have received payment in phase 2), and since we include the phase 3 dummy in all model specifications, the marginal effect of Post-Payment measures the change in contributions *in phase 3*, between those who received payment in phase 2 and those who did not. This finding confirms Hypothesis 4(i).

In model (V), even though the contribution subsidies are able to raise the average individual contribution by 5.35 in phase 2, they reduce the contributions by 1.30 in phase 3. Since the phase 3 dummy is not statistically significant, the average contributions of phases 1 and 3 are statistically not different. Thus, after controlling for all the covariates, the average contribution for groups that received payment in phase 2 decreased by 1.30 from phase 1 to phase 3, while the groups that did not received the payment experienced no statistically significant change in contributions between phases 1 and 3. This finding confirms Hypothesis 4(ii).

We have thus documented a case where "net crowding-out" occurs only in the long term: the payment subsidy worked to raise contributions when it is in place (in phase 2), but reduced contributions in the long term after it is over (in phase 3, and from phases 1 to 3). Deci (1971) observes a similar pattern for individual decisions where monetary interventions in period two reduced intrinsic motivation from periods one to three. His study focuses on private goods without externalities and does not consider the role of social norms. The long-term crowding out is in contrast to the child care case of Gneezy and Rusticini (2000), where the financial incentive of a fine did not work to encourage the desirable behavior even when the financial incentive was in place. Effectively, they showed that a strong short-term crowding-out effect of financial incentives can last until the incentive is over, while we show, in a group decision setting involving public goods, that long-term crowding-out can occur even when crowding-out does not occur in the short term.

The estimation results of Models (IV) and (V) show that the coefficients of most of the interaction terms involving Payment or Post-Payment are not statistically significant. The only exception is CDN*Payment, whose coefficient is negative and significant: for groups with strong collective norms, the introduction of contribution subsidies raises the contributions by a lesser amount, compared with groups with lower CDN. Figures 3 and 4 illustrate the marginal effects of Payment on average individual contributions for different values of CDN and PDN. In both cases, higher values of CDN or PDN reduce the impacts of Payment on contributions, but the role of PDN is not statistically significant. Figures 5 and 6 show the marginal effects of Post-Payment, again for different values of CDN and PDN. Both CDN and PDN tend to aggravate the negative impacts of Post-Payment: as CDN or PDN becomes larger, Post-Payment tends to reduce the average contributions further. However, the moderating effects of CDN and PDN are statistically insignificant.

In Model (V), we include Sorted and Sort-high, as well as their interactions with Payment and Post-Payment, in order to test Hypothesis 5. The direct effects of Sorted and Sort-high are statistically insignificant, indicating that, after controlling for all other covariates, homogeneous (i.e., sorted) groups behave similarly to the heterogeneous (i.e., unsorted) groups. However, the two kinds of groups tend to behave differently in responding to payment. Sorted groups tend to respond more to payment, although the coefficient of Sorted*Payment is not statistically significant. Sorted groups also reduce their contributions more in phase 3 after receiving payment in phase 2: the coefficient of Sorted*Post-Payment is negative and significant at the 10% level. That is, stopping the payment is more damaging to average contributions for homogeneous groups (i.e., sorted groups) compared with the more heterogeneous groups. Further, the difference is mostly between the Sort-low and Sort-medium groups on one hand, and unsorted groups on the other, as the positive coefficient of Sort-high*Post-Payment (2.20) more than compensates for the negative coefficient of Sorted*Post-Payment (-1.69). These results are consistent with the unconditional analysis in Table 3, which shows that the phases 1 to 3 reduction in contributions is the largest for Sort-low groups receiving payment.

Since we already control for the normative variables, the dummy Sorted might have captured additional factors that differ across homogeneous and heterogeneous groups. Alternatively, it is possible that the CDN and PDN measures do not completely capture the normative factors, and the residual effects are captured by Sorted. Nevertheless, the result shows that introducing and then stopping financial incentives might have larger negative impacts on collective action for homogeneous but not cooperative groups than for heterogeneous groups.

Evolution of normative perceptions

The long-term effects of contribution subsidies are also partly reflected by comparing the evolution of the perceived normative variables PDN and GI between payment and no-payment treatments. Figures 7 and 8 show the time patterns of average PDN and GI over the experiment periods under the payment and no-payment treatments, with each node representing the value at the *beginning* of the corresponding period.[6] PDN and GI are measured through subject surveys at the end of periods 1, 6, 12 and 18, with the values in other periods determined through linear extrapolation. Both figures demonstrate that *payment significantly raises PDN and GI*: although both perception variables decrease significantly in phase 1 from the end of Period 1 to that of Period 6 (graphed as starting values of Periods 2 and 7), contribution subsidies in phase 2 significantly raise both values by the end of Period 12 (which is the last period of phase 2). In contrast, PDN and GI values for groups that did not receive the contribution subsidies remain about unchanged in phase 2. In phase 3, once the contribution subsidies stop, PDN and GI values at the end of Period 18 effectively drop back to the levels found at the end of phase 1, wiping out the gains caused by the contribution subsidies in phase 2. On the other hand, the Figures show no evidence that experiencing Payment during phase 2 "destroys" PDN or GI, as their levels in phase 3 are comparable to those in phase 1.

Robustness checks

In Table 6, we present the results of a series of robustness checks. Model (i) repeats the estimates of Model (V) in Table 5. We remove the lagged dependent variable in (ii), and drop the first round in each phase, the last round, and both the first and last rounds in (iii) – (v). Including the lagged dependent

[6] Since CDN is defined as the lagged group average contribution in the previous period, its pattern is reflected in Figure 2, which shows the average contributions for all periods.

variable is intended to control for unobserved individual characteristics that are captured by the past contribution behavior. The first round might be particularly affected by the results in the previous phase, while the last round is special due to the dooms day effect. The results in Table 6 show that our estimates are extremely robust to these perturbations.[7] Further, while not reported in Table 6, the estimates are also robust to different ways of "filling in" the values of perception variables in (6), e.g., setting $\rho_{rj} = 0$ so that PDN and GI values in any round within a phase always equal to the measured value at the last round of the phase (when the PDN and GI survey questions are answered).

5. Conclusion

In this paper, we develop and test a conceptual model of the interactions of normative variables and financial incentives in influencing pro-social behavior, in order to understand the long-term consequences of short-term financial incentives. During the periods when the BPP is in effect, payments for contributions to the public good significantly raise the contribution levels and the payment's impact far outweighs those of the norms and group identification. However, all of the payment's direct effects on the contribution behavior dissipate after the payment is over. In fact, subjects tend to contribute less following the payment periods compared with those who do not receive payments at all, and the former even contributes less than before the payment was introduced. These findings indicate that even when payments do not completely crowd out other pro-social incentives in the short term (so that contributions increase in response to payments), they can have long-term crowding-out effects after the payment is over. Further, the long-term crowding-out effects vary across groups, with homogeneous and less cooperative groups experiencing the largest effect.

We find strong evidence that perceived descriptive norms and group identification interact to influence subject decisions on how much to contribute to the public good. Perceived descriptive norms raise the subjects' contributions by a larger amount when these subjects identify more with their groups compared with those who identify less with their groups. In contrast, collective descriptive norms do not have to interact with group identification: higher CDNs tend to raise a subject's contribution independent of her level of group identification.

These findings have two important implications. First, although economic research tends to focus only on observed collective decisions or norms, perceived norms can be important determinants of contribution behavior and are different from CDN. In a companion paper (Lapinski et al., 2016), we

[7] The only significant change occurs when the first round is dropped (Columns (iii) and (v)): the marginal effect of Payment goes down (to 3.35 and 3.06 respectively). This change indicates that payment affects the first round more than subsequent rounds, partly because in subsequent rounds, the conditioning variables such as CDN and lagged dependent variable already capture the effects of Payment.

study the evolution of PDN in detail, and show that CDN only has moderate influences on PDN.[8] More research is needed to study the role of PDN in broader economic and social decisions, and to study the relationship between PDN and CDN in broader settings. Second, relative to collective norms, understanding perceived norms is more important when promoting pro-social behavior in strongly identified groups. Since normative perceptions are closely related to behavioral decisions and PDNs can differ from CDNs, it is useful to understand what people perceive to be normative behavior. Likewise, out findings highlight the need, under some circumstances, to bring perceptions into line with reality (collective norms) through carefully designed communication interventions in order promote prosocial action. Importantly, the ethical aspects of this must be considered; such efforts (termed *normative restructuring*) should only be undertaken when social consensus exists on the behavior in question. For example, binge drinking of alcohol is widely agreed to be harmful making it a good candidate for such efforts (Perkins & Berkowitz, 1986).

Although short-term payments have long-term crowding-out effects, their impacts on perceived norms and group identification tend to be short term. Specifically, payment significantly raises the perceived descriptive norms and group identification, but all of these effects vanish after the payment is over. More research is needed to investigate the settings in which short-term payments may influence long-term norms. Overall, our findings send a strong message that accounting for both collective and perceived norms and bolstering connections with group members can enhance the effects of behavioral payment programs on pro-social behaviors over the long term.

References

Ariely, D., Bracha, A., and Meier, S. 2009. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, 99, 544-555.

Banerjee, A.V. 1992. A Simple Model of Herd Behavior. *Quarterly Journal of Economics* 107, 797–817.

Bikhchandani, S., Hirshleifer, D., and Welch I. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades. *Journal of Political Economy* 100, 992–1026.

Borsari, B., and Carey, K. B. 2003. Descriptive and injunctive norms in college drinking: A meta-analytic integration. *Journal of Studies on Alcohol*, *64*, 331–341.

Cameron, A.C. and Miller, D.L. 2015. A Practitioner’s Guide to Cluster- Robust Inference, *Journal of Human Resources*, 50, 317-372.

Cialdini, R.B., Reno, R.R., and Kallgren, C.A. 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of*

[8] Specifically, CDN’s major role is to mediate the influence of “Period 0 group average contribution” on PDN.

Personality and Social Psychology, 58, 1015-1026.

Deci, E.L., 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*, 105-115.

Deci, E.L., Koestner, R., Ryan, R.M., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125, 627-668.

DeJong, W., and Smith, S.W. 2013. Social norms marketing campaigns to reduce college student drinking: Truth in advertising. In R. Rice and C. K. Atkin (Eds.) *Public health campaigns* (3rd Ed., pp. 177-178), Thousand Oaks, CA: Sage.

Fehr E. and Schmidt, K.M. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114, 817-868.

Frey, B.S., Jegen, R., 2001. Motivation crowding theory. *Journal of Economic Surveys* 15, 589-611.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171-178. doi:10.1007/s10683-006-9159-4

Gneezy, U., & Rustichini, A. (2000). Fine is a price, a. *J. Legal Stud.*, *29*, 1.

Goeschl, T., Perino, G., 2012. Instrument Choice and Motivation: Evidence from a Climate Change Experiment. *Environmental and Resource Economics* 52(2), 195-212.

Hollander, H., 1990. A social exchange approach to voluntary cooperation. *American Economic Review* 80, 1157-1167.

Hunter, J. E., and Gerbing, D. W. 1982. Unidimensional measurement, second order factor analysis, and causal models. *Research in organizational behavior, 4*, 267-320.

Kits, G. J., Adamowicz, W. L., & Boxall, P. C. (2014). Do conservation auctions crowd out voluntary environmentally friendly activities? *Ecological Economics*, *105*, 118–123. https://doi.org/10.1016/j.ecolecon.2014.05.014

Lapinski, M. K., Maloney, E. K., Braz, M. E., & Shulman, H. C. 2013. Using normative messages to improve hand washing: A field experiment. *Communication Research, 39,* 21-46.

Lapinski, M., & Rimal, R. 2005. An explication of social norms. *Communication Theory* 15: 127-147. doi:10.1111/j.1468-2885.2005.tb00329.x

Lapinski, M. K., Zhuang, J., Koh, H., Shi, J. Y. 2015. Social norms and involvement in health and environmental behaviors. *Communication Research*, 1-24. DOI: 10.1177/0093650215605153

Lapinski, M. K., Kerr, J. M., Zhao J., Shupp R. S., 2016. Social Norms, Behavioral Payment Programs, and Cooperative Behaviors: Toward a Theory of Financial Incentives in Normative Systems. *Human Communication Research*, http://dx.doi.org/10.1111/hcre.12099

Levitt, S.D., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21, 153-174.

Meier, S. 2007. “A Survey on Economic Theories and Field Evidence on Pro-social Behavior.” In conomics and Psychology: A Promising New Cross-Disciplinary Field, ed. Bruno S. Frey and Alois Stutzer, 51–88. Cambridge: MIT Press.

Mollen, S., Rimal, R. N., & Lapinski, M. 2010. What is normative in *Health Communication* research on norms? A review and recommendations for future scholarship. *Health Communication*, *25*(6-7), 544-547.

Perkins, H. W. and Berkowitz, A. D. 1986. Perceiving the community norms of alcohol use among students: some research implications for campus alcohol education programming. *Substance Use & Misuse, 21*, 961-976.

Rimal, R. 2008. Modeling the relationship between descriptive norms and behaviors: A test and extension of the theory of normative social behavior (TNSB). *Health Communication, 23*, 103-116. doi: 10.1080/1041023080196779

Rimal, R., and Real, K. 2005. How behaviors are influenced by perceived norms. *Communication Research, 32,* 389-414. doi: 10.1177/0093650205275385

Rode, J., Gómez-Baggethun, E., and Krause, T. 2015. Motivation crowding by economic incentives in conservation policy: A review of the empirical evidence. *Ecological Economics*, *117*, 270-282.

Sethi, R., Somanathan, E. 1996. The evolution of social norms in common property resource use. *American Economic Review* 86: 766–788.

Smith, S.W., Atkin, C., Martell, D., Allen, R., & Hembroff, L. 2006. A social judgment theory approach to conducting formative research in a social norms campaign, *Communication Theory*, 16: 141-152.

Vollan, B., 2008. Socio-ecological explanations for crowding-out effects from economic field experiments in Southern Africa. *Ecological Economics* 67, 560-573.

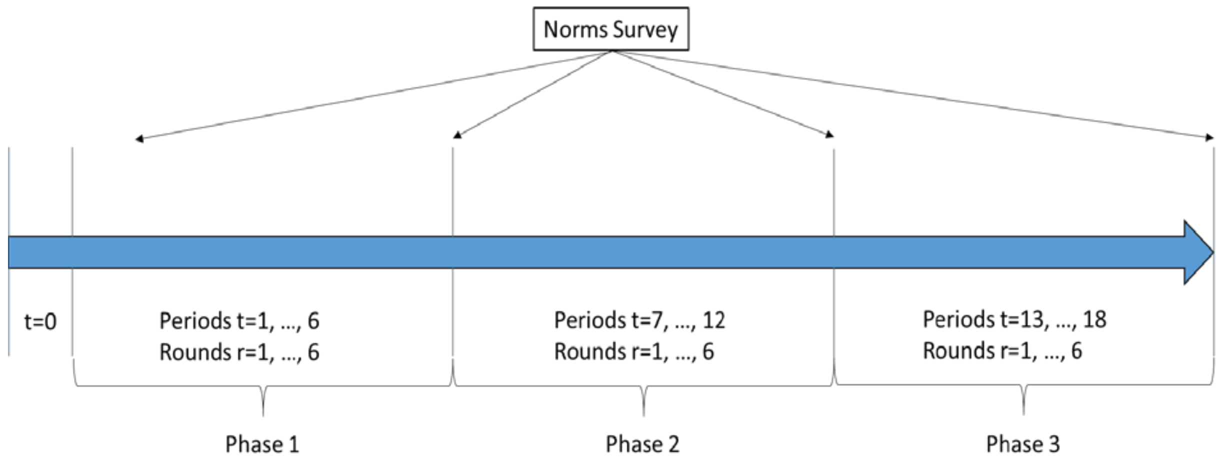

Figure 1. Experiment Procedure: Periods, Rounds and phases

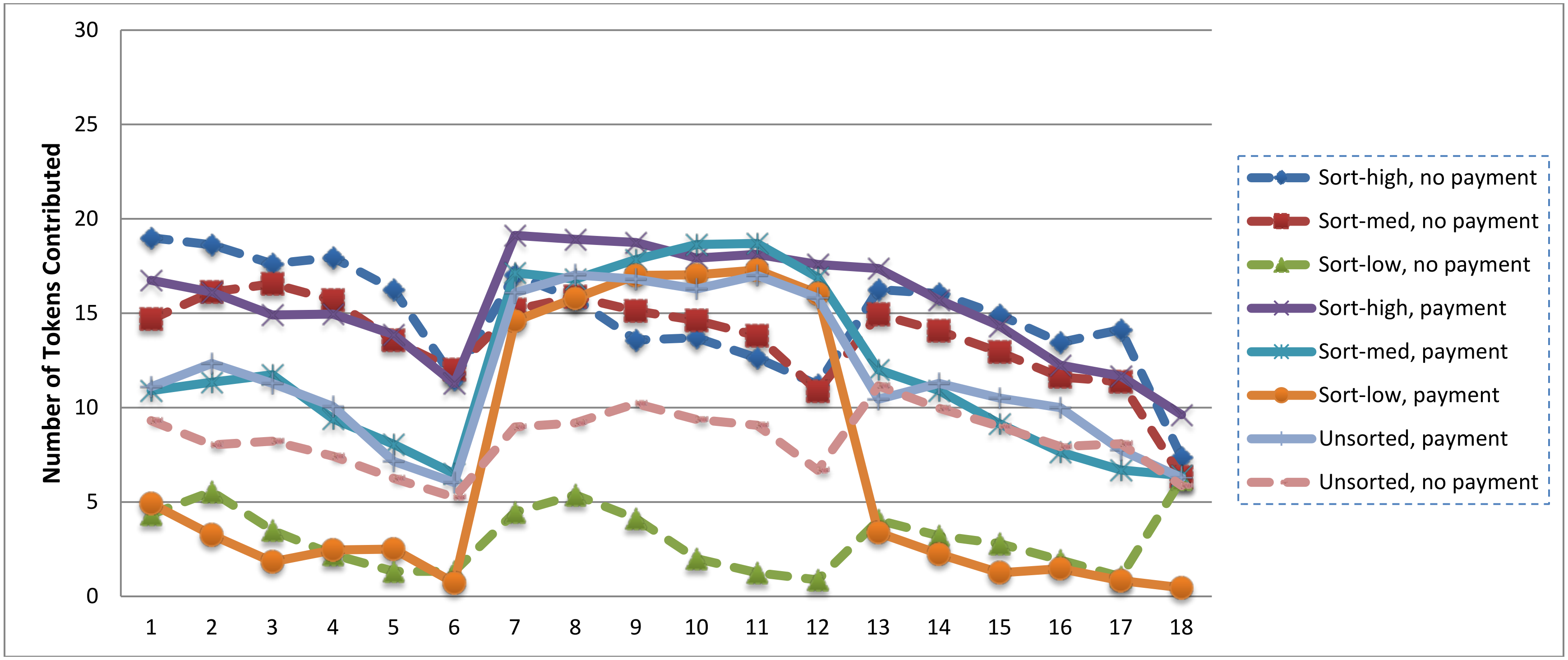

Figure 2. Average contribution by period, for payment and sorting treatments

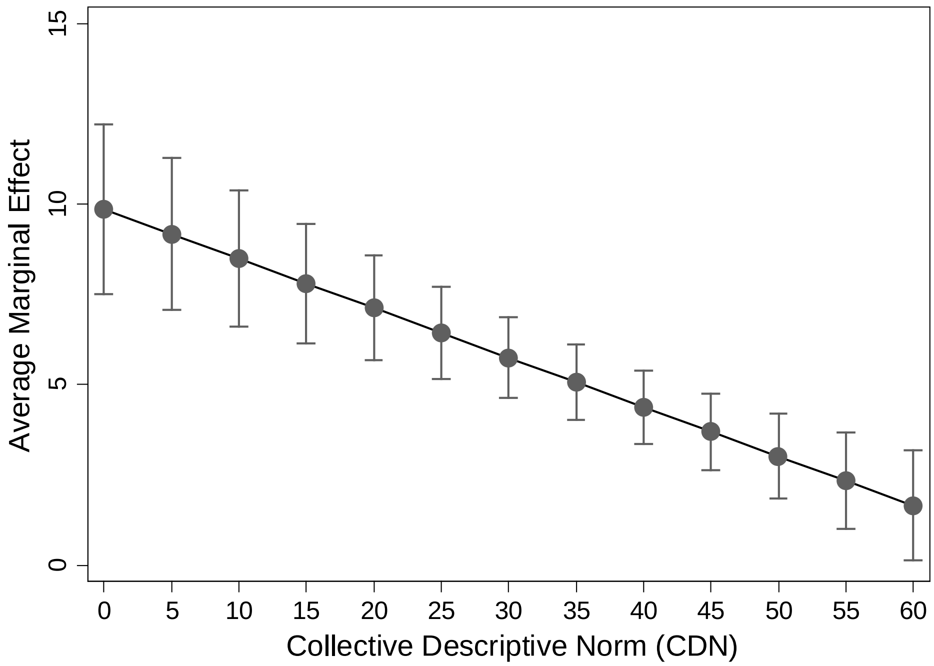

Figure 3. Average marginal effects of Payment on Contributions conditional on CDN, with 95% confidence intervals

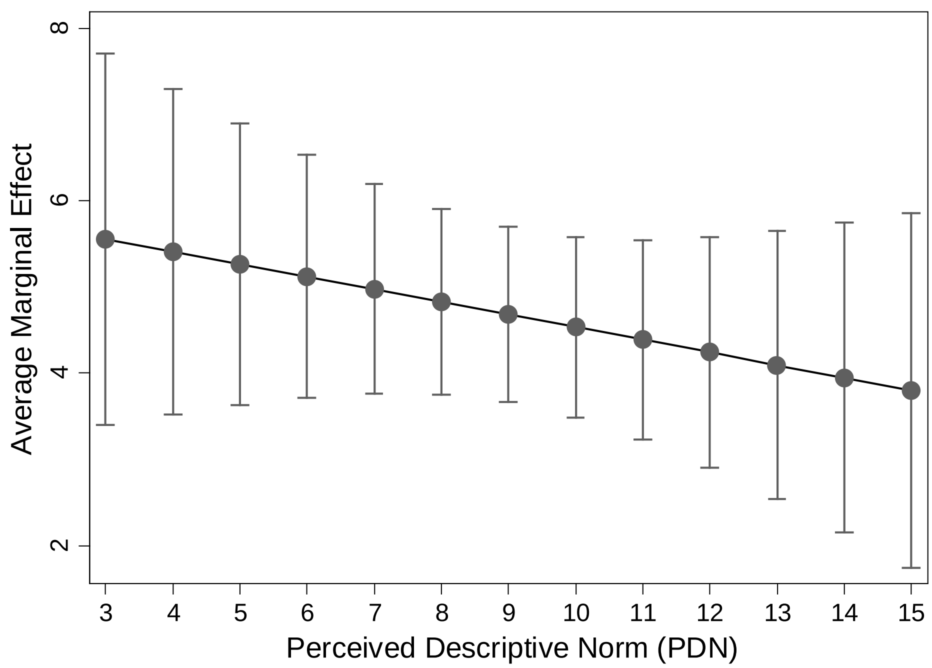

Figure 4. Average marginal effects of Payment on Contributions conditional on PDN, with 95% confidence intervals

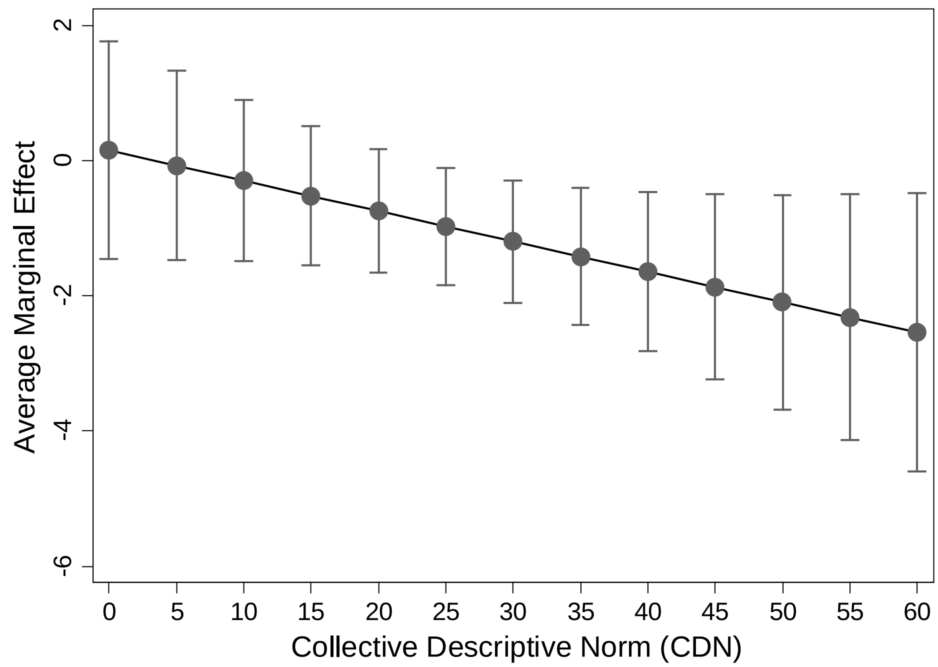

Figure 5. Average marginal effects of Post-Payment on Contributions, conditional on CDN, with 95% confidence intervals

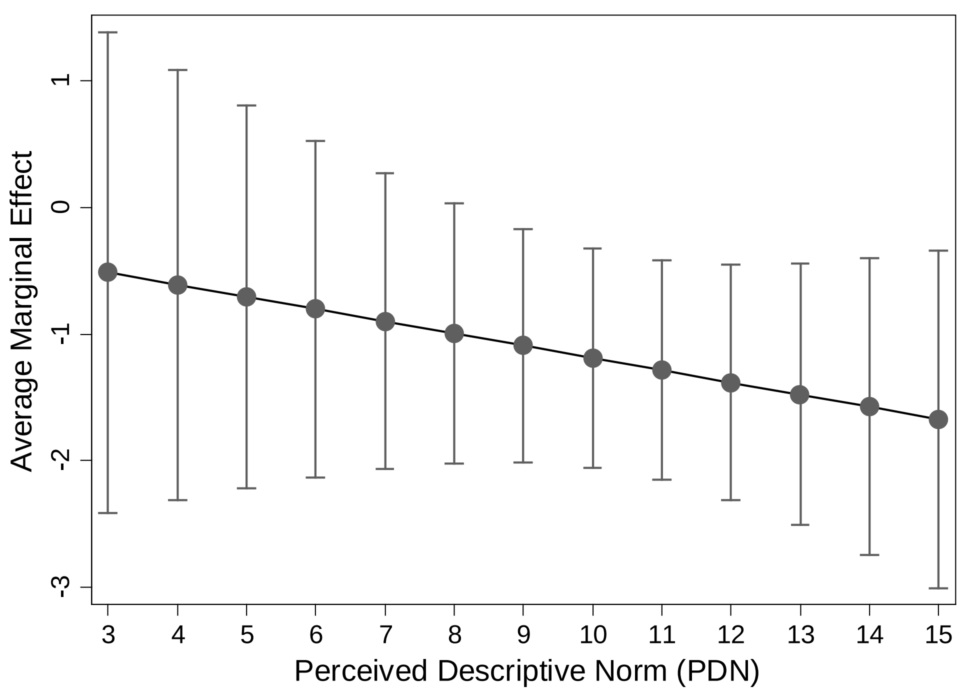

Figure 6. Average marginal effects of Post-Payment on Contributions conditional on PDN, with 95% confidence intervals

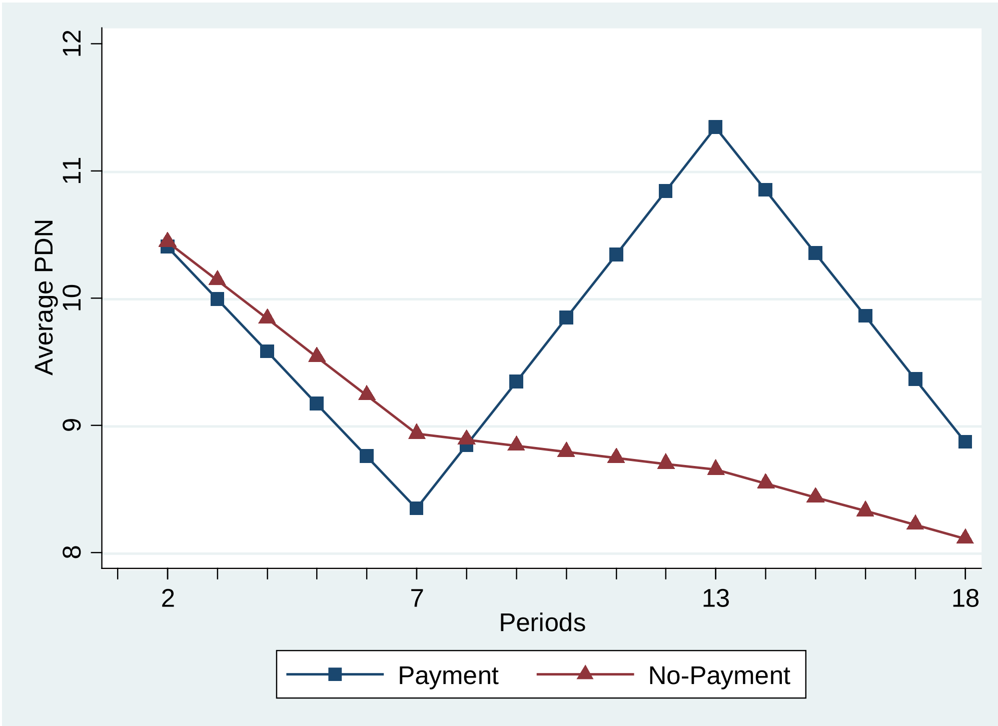

Figure 7. Evolution of Average PDN, with and without Payment

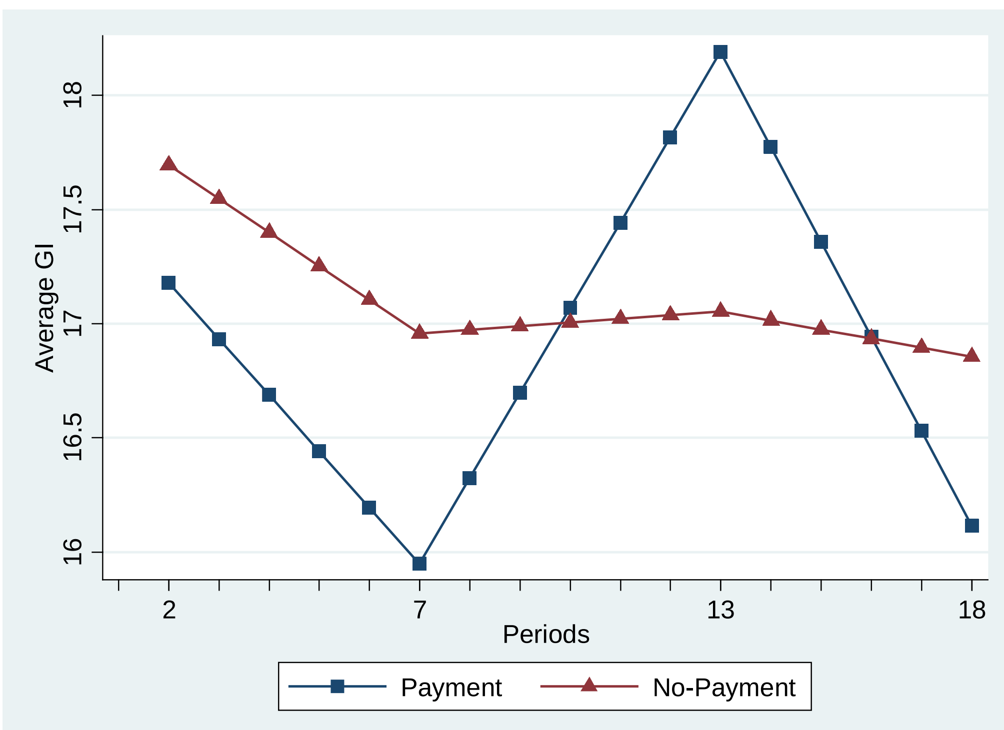

Figure 8. Evolution of Average GI, with and without Payment

Table 1. Sample sizes in different treatments

	Incentive Payment	No Incentive
Sorted	68 subjects; 3 sessions (24, 24, 20); 17 groups	64 subjects; 3 sessions (24, 20, 20); 16 groups
Not Sorted	32 subjects; 2 sessions (16, 16); 8 groups	28 subjects; 2 sessions (16, 12); 7 groups

Note: Each cell corresponds to a particular treatment. It describes the total number of subjects in the particular treatment; the number of sessions conducted, with numbers in brackets describing the number of subjects in each session; and the number of subject groups, with each group consisting of four subjects playing the public good game.

Table 2. Survey questions about PDN and group identification

Questions about PDN

The following questions ask for your thoughts about your group members in this game. Please circle the number that corresponds to the response that best describes your thoughts about them.

A. I think that most people in my group are cooperative.
□□□□□ Strongly Disagree □ Disagree □ Neutral □ Agree □ Strongly Agree
B. The majority of people in my group are playing this game cooperatively.
□□□□□ Strongly Disagree □ Disagree □ Neutral □ Agree □ Strongly Agree
C. It seems like most people in this group do not cooperate with each other.
□□□□□ Strongly Disagree □ Disagree □ Neutral □ Agree □ Strongly Agree

Questions about Group Identification

We would like to know a few things about you. Please think about people in this group for these questions. On the whole, how similar do you think most people in this group are to you?

	Not at all similar					**Extremely similar**	
1. intellectually?	1	2	3	4	5	6	7
2. in the way they think?	1	2	3	4	5	6	7
3. in their values?	1	2	3	4	5	6	7
4. in their behaviors?	1	2	3	4	5	6	7

Table 3. Average contributions for payment and sorting treatments

		Payment				No Payment			
		Overall	Sort-high	Sort-medium	Sort-low	Overall	Sort-high	Sort-medium	Sort-low
Sorted	Initial (t=0)	10.43 (7.06)	17.67 (3.05)	10.00 (3.53)	2.25 (3.65)	11.33 (6.74)	18.21 (2.58)	10.04 (3.44)	2.94 (3.18)
	phase 1	8.93 (7.75)	14.65 (7.22)	8.18 (5.77)	2.21 (4.17)	11.14 (8.04)	16.83 (5.58)	11.36 (6.76)	2.29 (4.15)
	phase 2	17.44 (5.04)	18.40 (3.89)	17.76 (4.08)	15.92 (6.71)	9.94 (7.99)	13.99 (7.32)	11.14 (6.80)	2.06 (4.17)
	phase 3	7.92 (7.84)	13.49 (7.90)	7.70 (6.03)	1.48 (3.50)	9.30 (8.02)	14.35 (6.74)	9.17 (7.29)	1.94 (4.19)
	% reduction from phases 1 to 3	11%	8%	6%	33%	17%	15%	19%	15%
Not sorted	Initial (t=0)	9.72 (6.00)				9.93 (7.68)			
	phase 1	9.66 (6.90)				7.98 (8.00)			
	phase 2	16.51 (5.74)				8.83 (7.58)			
	phase 3	9.38 (7.53)				7.70 (7.89)			
	% reduction from phases 1 to 3	3%				4%			

Note: Standard deviations in parentheses.

Table 4. Explanation of Variables and Summary Statistics

Variable	Explanation	N	Mean	Std. Dev.	Min	Max
a_{nkrj}	Public good contribution of individual *n* of group *k* in round *k* of phase *j*	3,456	10.59	8.07	0	20
$a_{n,0}$	Period 0 contribution	3,456	10.54	6.91	0	20
$g_{k,0}$	Period 0 group average contribution	3,456	10.54	5.84	0	20
$Payment_{nkrj}$	Dummy for contribution subsidy (=1 for subsidy rounds and =0 otherwise)	3,456	0.17	0.38	0	1
CDN_{nkrj}	Collective descriptive norm	3,456	32.56	20.16	0	60
PDN_{nkrj}	Perceived descriptive norm	3,072	9.44	2.96	3	15
GI_{nkrj}	Group similarity	3,072	17.04	5.11	4	28
$Sorted_k$	Dummy for sorted groups	3,456	0.69	0.46	0	1
$Sort\text{-}High_k$	Dummy for sort-high groups	3,456	0.25	0.43	0	1

Table 5. Estimation Results: Individual Contribution Decisions

	Dependent variable: contribution level				
	(I)	(II)	(III)	(IV)	(V)
Period 0 contribution	0.08	0.12^{*}	0.11^{*}	0.12^{*}	0.17^{***}
	(0.05)	(0.07)	(0.06)	(0.07)	(0.04)
Period 0 group ave. contribution	0.22^{***}	0.11	0.11	0.08	
	(0.07)	(0.08)	(0.08)	(0.08)	
Payment	4.67^{***}	4.13^{***}	4.19^{***}	11.98^{***}	10.82^{***}
	(0.65)	(0.61)	(0.61)	(1.39)	(1.41)
Post-Payment	-0.86	-1.01^{**}	-1.08^{**}	-0.04	1.69
	(0.57)	(0.49)	(0.49)	(1.15)	(1.28)
CDN		0.05^{***}	0.06^{*}	0.08^{***}	0.09^{***}
		(0.01)	(0.03)	(0.03)	(0.03)
PDN		-0.02	-0.44^{***}	-0.29^{*}	-0.27^{*}
		(0.06)	(0.16)	(0.14)	(0.14)
GI		0.04	-0.17^{**}	-0.11^{*}	-0.11^{*}
		(0.03)	(0.07)	(0.06)	(0.06)
CDN*GI			-0.00	-0.00	-0.00
			(0.00)	(0.00)	(0.00)
PDN*GI			0.03^{***}	0.02^{**}	0.02^{**}
			(0.01)	(0.01)	(0.01)
CDN*Payment				-0.13^{***}	-0.14^{***}
				(0.03)	(0.03)
CDN*Post-Payment				-0.02	-0.04
				(0.02)	(0.03)
PDN*Payment				-0.23	-0.15
				(0.15)	(0.15)
PDN*Post-Payment				-0.04	-0.10
				(0.13)	(0.11)
Sorted					0.03
					(0.43)
Sort-high					-0.03
					(0.63)
Sorted*Payment					0.88
					(0.78)

Sort-high*Payment					-0.94 (0.94)
Sort*Post-Payment					-1.69^{*} (0.87)
Sort-high*Post-Payment					2.20^{*} (1.30)
Phase2	0.21 (0.41)	0.22 (0.32)	0.25 (0.33)	0.47^{*} (0.27)	0.48^{*} (0.27)
Phase3	-0.23 (0.39)	-0.16 (0.30)	-0.11 (0.30)	0.17 (0.24)	0.20 (0.24)
Lagged Dependent Variable	YES	YES	YES	YES	YES
Round dummy	YES	YES	YES	YES	YES
N	3456	3072	3072	3072	3072
r2	0.61	0.61	0.61	0.63	0.63
Marginal Effects					
CDN		0.05^{***} (0.01)	0.05^{***} (0.01)	0.05^{***} (0.01)	0.05^{***} (0.01)
PDN		-0.02 (0.06)	-0.01 (0.06)	-0.04 (0.06)	-0.02 (0.05)
Payment (in phase 2)	4.67^{***} (0.65)	4.13^{***} (0.61)	4.19^{***} (0.61)	5.45^{***} (0.52)	5.35^{***} (0.54)
Post-Payment (in phase 3)	-0.86 (0.57)	-1.01^{**} (0.49)	-1.08^{**} (0.49)	-1.24^{**} (0.49)	-1.30^{**} (0.47)

Clustered (by group) robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All estimation equations include lagged individual contribution and dummies for rounds (2-6) and phases (2,3).

Table 6. Robustness Check: Individual Contribution Decisions

	(i) Full Sample	(ii) No lagged dependent variable	(iii) First round dropped	(iv) Last round dropped	(v) Both first and last rounds dropped
Period 0 contribution	0.17^{***} (0.04)	0.29^{***} (0.08)	0.07^{*} (0.04)	0.19^{***} (0.04)	0.08^{**} (0.04)
Payment	10.82^{***} (1.41)	13.08^{***} (1.78)	6.23^{***} (1.58)	11.45^{***} (1.31)	6.29^{***} (1.35)
Post-Payment	1.69 (1.28)	2.91 (1.98)	2.26^{***} (0.79)	0.87 (1.58)	1.67^{*} (0.97)
CDN	0.09^{***} (0.03)	0.19^{***} (0.04)	0.10^{***} (0.03)	0.11^{***} (0.03)	0.12^{***} (0.03)
PDN	-0.27^{*} (0.14)	-0.59^{**} (0.24)	-0.35^{***} (0.12)	-0.22 (0.15)	-0.32^{**} (0.13)
GI	-0.11^{*} (0.06)	-0.23^{**} (0.10)	-0.15^{***} (0.05)	-0.07 (0.06)	-0.13^{**} (0.06)
CDN*GI	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
PDN*GI	0.02^{**} (0.01)	0.04^{***} (0.01)	0.02^{***} (0.01)	0.02^{*} (0.01)	0.02^{**} (0.01)
CDN*Payment	-0.14^{***} (0.03)	-0.12^{***} (0.03)	-0.07^{**} (0.03)	-0.15^{***} (0.03)	-0.08^{***} (0.03)
CDN*Post-Payment	-0.04 (0.03)	-0.03 (0.03)	0.00 (0.02)	-0.05^{*} (0.03)	-0.01 (0.02)
PDN*Payment	-0.15 (0.15)	-0.36 (0.21)	-0.07 (0.11)	-0.19 (0.16)	-0.07 (0.11)
PDN*Post-Payment	-0.10 (0.11)	-0.22 (0.16)	-0.18^{**} (0.08)	-0.03 (0.13)	-0.10 (0.09)
Sorted	0.03 (0.43)	0.03 (0.68)	0.15 (0.37)	-0.07 (0.54)	0.10 (0.41)
Sort-high	-0.03 (0.63)	-0.08 (0.99)	-0.02 (0.52)	-0.05 (0.68)	-0.08 (0.45)

Sorted*Payment	0.88	1.05	0.36	1.04	0.50
	(0.78)	(0.99)	(0.62)	(0.81)	(0.61)
Sort-high*Payment	-0.94	-1.46	-0.53	-1.09	-0.76
	(0.94)	(1.26)	(0.70)	(0.99)	(0.65)
Sort*Post-Payment	-1.69*	-1.92	-0.81	-1.77*	-0.92
	(0.87)	(1.15)	(0.52)	(1.02)	(0.61)
Sort-high*Post-Payment	2.20*	2.04	0.06	2.38	0.15
	(1.30)	(1.64)	(0.89)	(1.42)	(1.10)
Phase2	0.48*	0.38	0.18	0.62*	0.30
	(0.27)	(0.40)	(0.19)	(0.36)	(0.29)
Phase3	0.20	-0.01	-0.14	0.37	0.02
	(0.24)	(0.40)	(0.20)	(0.31)	(0.22)
Lagged Contribution	0.46***		0.56***	0.45***	0.57***
	(0.03)		(0.03)	(0.03)	(0.03)
Round dummy	YES	YES	YES	YES	YES
N	3072	3072	2688	2496	2112
r2	0.63	0.53	0.68	0.64	0.70
Marginal Effects					
CDN	0.05***	0.15***	0.08***	0.05***	0.09***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
PDN	-0.02	-0.07	-0.05	0.00	-0.03
	(0.05)	(0.08)	(0.04)	(0.06)	(0.05)
Payment	5.35***	6.27***	3.35***	5.16***	3.06***
	(0.54)	(0.62)	(0.61)	(0.56)	(0.55)
Post-Payment	-1.30**	-0.92	0.16	-1.76***	-0.19
	(0.47)	(0.65)	(0.29)	(0.53)	(0.27)

Clustered (by group) robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$